V.80 V 84 V95

SOME METHODOLOGICAL INVESTIGATIONS ON PRE-HARVEST FORECASTING OF JUTE CROP BASED ON BIOMETRICAL OBSERVATIONS

B.S. RAMESH

Dissertation submitted in fulfilment of the requirements for the award of Diploma in Agricultural Statistics of the Institute of Agricultural Research Statistics (I.C.A.R.), New Delhi-12

DE04/9

INSTITUTE OF AGRICULTURAL RESEARCH STATISTICS
(I.C.A.R.)
NEW DELHI
1 9 7 4

ACKNOWLEDGEMENTS

I feel great pleasure in expressing my deep sense of gratitude to Shri K.S. Krishnan, Senior Statistician, Institute of Agricultural Research Statistics (I.C.A.R.), far his inspiring guidance, keen interest and constant encouragement throughout the course of investigation and preparation of this thesis.

I am highly thankful to Shri M.P. Jha, Statistician-cum-Associate Professor, I.A.R.S., for making available the detailed data used for the study and continuous help in the development of the analytical procedure and for his valuable suggestions and encouragement throughout the course of study.

I am also thankful to Dr. D. Singh, Director, I.A.R.S., for providing the adequate facilities to carry out the research work.

The financial assistance provided in the form of Fellowships by the Indian Council of Agricultural Research is hereby gratefully acknowledged.

B. S. Ramesh.

(B.S. RAMESH)

I.A.R.S.(I.C.A.R.)

NEW DELHI-110012

CONTENTS

CHAPTER - I

INTRODUCTION

The growing importance of commercial crops has increased the need for accurate forecasts and estimates of its production. Hence reliable forecasts of crop production before the harvest will be of great help to the trade, Industry and to the Government in making its price and export-import policies.

Among the various branches of economic activity, agriculture is the only one which seems to be subject to wide and irregular fluctuations of output. Whereas fluctuations of Industrial production can be attributed almost entirely to social factors which are amenable to control, the year-to-year variations of agricultural output are largely determined by physical factors. Reliable crop forecasts which may allow the necessary social adjustments to be made in advance are therefore of particular importance.

Of the two components of crop production, acreage and yie acreage presents a comparatively simpler problem. Ordinarily, acreage is much more stable than yield per acre. For certain purposes, it is desirable to forecast acreage in advance of planting. Since the acreage of a cash crop is usually influenced by the ratio of the price of that crop to those of competitive crops which prevailed during the previous year (or years), forecasts of acreage planted have often been based on a regression analysis of these relationships. In addition, weather factors may sometimes be used to predict changes in acreage.

The possibility of forecasting crop yields several years in advance would, of course, be of great value in the planning of agricultural production. There are three principal approaches to forecasting crop yields.

(a) Forecasts based on the observations on crops environment such as weather and input supplies etc;

(b) Forecasts based on eye observations of the growing crop;

(c) Forecasts based on measurements of the growing crop in typical fields.

These approaches are complementary and may be combined. Such forecasts which are based on reports of farmers or officials are to a greater degree subject to vicissitude in human judgement. For example, farmer's appraisals tend to be conservative in a year following a poor crop. Hence these approaches were found to be subjective. It needs to be investigated whether the forecasting of crop production could be made more reliable by adopting objective methods such as by taking into consideration the biometrical observations on the crop during its various growth stages.

<u>Objective Methods:</u> These methods are being developed under two broad lines viz;

(a) Methods based on objective sample measurements of the growing crop. The problem here is to find some plant characteristics which can be measured well in advance of the harvest and which can be used as indicators of final yield.

(b) Methods based on observations of the environment of the crop, particularly weather factors.

Frequently a combination of these two approahhes prove superior to either method used alone. Since for our present study data on weather factors for small units were not available, we confine our methodology to the first approach only.

<u>Advatages of pre-harvest forecasting of crop yields based on plant counts and measurements:</u>

The forecasts of crop yields from plant counts and measurements, in lieu of farmers or officials reports on crop production, has at least two major advantages.

(i) By-product information that is available or obtainable by making minor modifications; and

(ii) Greater objectivity

Useful by-product information which can be conveniently collected includes changes in components or attributes of yield over time and comparisons of yield characteristics among varieities or cultural practices. With regard to objectivity, farmers or officials appraisal may not include an accurate reflection of the impact of recent changes in varieties or cultural practices. Although changing farm practices may alter the parameters in the models, ane may feel confident that a forecasting system based on plant counts and measurements is more responsive to changes in farm practices than farmer's appraisals.

*   *In some countries post harvest estimates are worked out on the basis of farmers reports. However, estimates derived from preharvest sampling are more useful since these are objective and will be available earlier than estimates from post-harvest farmer's report. Preharvest sampling also provides a means of getting much valuable information that cannot otherwise be easily obtained. By means of laboratory analysis of samples taken from fields, information on various attributes of crop quality can be made available. Also, if deemed worthwhile, information on some types of insect damage, can be readily obtained.

<u>Scope of the present study</u>: The present study pertains to the methodology of preharvest forecasting of Jute crop in West Bengal during the years 1970-71, 71-72, 72-73 and 73-74. Jute is one of the most important cash crop in India and ranks high among the commercial crops grown in the country. Its cultivation is almost entirely confined to the Eastern states of West Bengal, Assam, Bihar, Orissa and Tripura. The main reasons for the concentration of the cultivation of this fibre crop in these states are the favourable soil and weather conditions, availability of cheap and skilled labour

---

* F.A.O. of United Nations

and retting facilities. The Jute textile industry is one of the major industries in India, yielding substantial foreign exchange through export of its manufactures.

*The area under Jute crop in the country was 5.71 lakh hectares during the year 1950-51. Subsequently it showed considerable annual fluctuations. During the years 1969-70, 70-71 and 71-72, it was 7.68, 7.50 and 8.15 lakh hectares respectively. The productionof Jute crop for the years 1969-70,70-71 and 71-72 was 5.66, 4.94 and 5.71 million bales respectively. The likely achievement in 1973-74 is of the order of 6.00 to 6.20 million bales. The production targets and achievements in the industrial sector of Jute manufactures (in thousand tonnes unit) is as follows:

1969-70	(Actual achievement)	944
1970-71	(Actual achievement)	1050
1971-72	(Anticipated)	1100
1973-74	(Target)	1400
1973-74	(Likely achievement)	1400

The importance of Jute industry, as a source of foreign exchange is reflected in the fact that during 1969-70 the value of Jute goods exported from India was Rs.205 crores against a total export trade valued at Rs.1413 crores. It amounts to nearly 15% of the total trade.

*Ministry of Food and Agriculture, Government of India.
Planning Commission, Government of India.

Even though the system of crop forecasting was enforced in India as early as in 1884, the Jute crop was included in this system only in 1943-44. The Directorate of Economics and Statistics, Ministry of Agriculture, Government of India issues estimates of Jute and other crops on the basis of information received from different states. In case of Jute crop, two forecasts are issued, the first one by the end of second week of August (about three months after sowing) and the second and the final one by the first week of November.

The first forecast provides information regarding the areas sown, germination and prevailing weather conditions. This forecast relates to acreage and does not deal with advance information on production. The second and final forecast is based on complete enumeration of all the fields under the crop. Production estimates in these forecasts are based on the partial reports of the crop cutting experiments conducted on early harvested fields. Final estimates of Jute production are based on complete results of crop cutting experiments. Except for this final estimate, which is really a post-harvest forecast, the forecasts of production issued are not based on any scientific methods and are therefore not of assured reliability.

With a view to bring about improvement in the forecasting of Jute crop through the use of measurements on biometrical characters in randomly selected fiel ds, the Institute of Agricultural Research Statistics, in collaboration with the Department of Agriculture, West Bengal and the

Jute Agricultural Research Institute, Barrackpore, undertook a series of pilot studies on Jute crop.

The above mentioned pilot studies involves important statistical problems in the analysis of the data collected. Some of the problems discussed in the present work are as follows:

(1) Choice of appropriate model to study the regression of yield on bimetrical characters using the classical least squares principle.

(2) Appropriate level, such as districts or blocks, at which the prediction equations can be constructed and used.

(3) Estimates of true variance components for between circles, between fields within circles and between plots within fields.

(4) Forecasting current and future years crop yields by making use of equations pertaining to two or more of the preceding years.

CHAPTER - II

MATERIAL AND METHODS OF ANALYSIS

2.1 Material: The present study deals with the data on Jute crop collected in the districts of Nadia and 24-parganas (N) of West Bengal during the years 1970-71 to 1973-74. The field survey work was carried out in the special package programme areas of the Jute crop in the blocks (i) Bhimpur (Nadia), (ii) Haringhatta (Nadia), (iii) Baduria (24-parganas) and (iv) J.A.R.I. Extension (24-parganas) during each of the four years mentioned above.

The variety of the crop in the areas covered under the surveys was generally olitorious. This variety is sown during the period April-May and harvested during August-October. The soil type of these areas is broadly alluvial and the average rainfall is about 120 Cm.

Organisation of the survey work: The field work of the surveys was carried out in two blocks in each of the two districts viz., Nadia and 24-parganas (N) by the Development Staff (Jute Field Assistant/Village Level Worker) of the State Department of Agriculture in addition to their normal duties. The number of persons deputed for the survey work in each block was as under:

	Block	No. of persons deputed
(i)	Bhimpur (Nadia)	16
(ii)	Haringhatta (Nadia)	14
(iii)	Baduria (24-parganas-North)	12
(iv)	J.A.R.I. Extn. (24-parganas-North)	6
	TOTAL:	48

The field staff of the Regional Office Jute Development, Government of India, helped the above mentioned development staff of the Department of Agriculture, West Bengal in the collection of the required field data.

In order to have effective supervision of the field work, full time supervisory staff, consisting of one field officer and four field supervisors, were provided. Besides the supervisory staff appointed under the project, the field work was also supervised by the officers of the State Department of Agriculture and the staff of J.A.R.I., Barrackpore and I.A.R.S., Delhi were connected with the project. The field staff were thoroughly trained in the collection of biometrical observations before starting the field work in each year. The pattern of the staff engaged in the survey work remained almost similar from year to year.

The sampling design and recording of observations: For recording the data, two stage stratified random sampling design was adopted. Four community development blocks covered under the special package programme, were taken for the study each year. The number of V.L.W./J.F.A. circles in the blocks were varying from six to seventeen. These circles constituted the strata. For each circle a list of all the fields under the special Jute package programme was prepared and a random sample of five fields was selected from each such list. In each selected field two plots each of size 2m x 2m were selected at random for recording detailed biometrical observations.

According to the plan of the work, the biometrical characters were recorded at various growth stages of the Jute crop. The first observations were recorded a month after the sowing of the crop followed by periodic observations at intervals of four weeks upto and including the time of harvest.

The total number of plants were counted in each selected plot on each occasion for recording the observations. For measuring the plant height and basal diameter during the growth period, five plants, four corner plants and one central plant in each of the selected plot were taken into consideration. The fibre weight of Jute was taken after retting the crop. Some data were rejected from the list of fields in the programme of the study due to non-availability of data relating to fibre weight after retting and due to partial recording of the biometrical observations.

2.2 Methods of Analysis

(a) Using the principle of Least Squares the regression of fibre weight of Jute on plant density, plant height and basal diameter has been studied under the following models;

$$\text{(i)}\quad Y = a_1 + b_1 x_1 + c_1 x_2 + d_1 x_3$$

$$\text{(ii)}\quad \text{Log } Y = a_2 + b_2 \text{ Log } x_1 + c_2 \text{ Log } x_2 + d_2 \text{ Log } x_3$$

where x_1, x_2, x_3 and y denote the plant density, plant height, basal diameter and expected fibre yield corresponding to a set of values of x_1, x_2 and x_3.

Keeping the dependent variables same, in the above mentioned two models and then applying different scalar transformations to the three independent variables viz., linear, logarithmic, square root and reciprocal, the relative prediction value of the equations were tested by working out the values of R, the multiple correlation coefficient and the corresponding residual sum of squares. We choose that equation which yields minimum residual sum of squares. It can be easily seen that the number of combinations for each model is 64 if each of the three independent variables is transformed in turn with such transformations of the other two variables. The 64 equations are directly comparable since the dependent variable and the number of parameters to be estimated are same for the equations.

For the choice between regression equations based on linear and loglinear transformations of the dependent variables, we cannot apply the technique of minimum residual sum of squares, because the difficulty arises due to scaling factor. The variance of Y changes with the units of measurements of Y, but the variance of log Y does not, because $\log cY = \text{Log } c + \text{Log } Y$ and the addition of a constant (log c) does not alter the variance. A direct comparison of residual sum of squares is therefore meaningless because by a proper choice of units of measurements one residual sum of squares may be made smaller than the other.

By standardizing the variable Y in such a way that its variance does not change with units of measurement we may bring these two equations onto a common footing. If we do the transformation so that the 'Jacobian' of transformation is the same for Y^* and Log Y^*, where Y^* is transformed Y, we can directly compare the residual sum of squares.

A transformation of Y that allows such a comparison of the residual sum of squares may be defined as

$$Y^* = (c) x (Y)$$

Where

$$C = \text{exponential} \left(- \frac{\sum \text{Log } Y}{N} \right)$$

is the inverse of the geometric mean of Y.

By standardizing Y by its geometric mean and defining the standardized value of Y^* we may express the two equations (i) and (ii) in terms of Y^* rather than Y as

$$\text{(iii)} \quad Y^* = \overset{*}{a}_1 + \overset{*}{b}_1 x_1 + \overset{*}{c}_1 x_2 + \overset{*}{d}_1 x_3$$

$$\text{(iv)} \quad \text{Log } Y^* = \overset{*}{a}_2 + \overset{*}{b}_2 \text{ Log } x_1 + \overset{*}{c}_2 \text{ Log } x_2 + \overset{*}{d}_2 \text{ Log } x_3$$

Since the residual sum of squares in these two equations, (iii) and (iv) are directly comparable, we choosee the functional form yielding the minimum residual sum of squares as the empirically appropriate functional form.

A non-parametric test is used to see whether the difference between the residual sum of squares in these two functional forms is significant. The test is based on a statistic defined as

$$d = \frac{N}{2} \left| \text{Log} \frac{\sum e_1^2}{\sum e_2^2} \right|$$

Where $\sum e_1^2$ and $\sum e_2^2$ are the residual sums of squares in estimating equations (iii) and (iv) respectively. They are based on the same number of degrees of freedom. The d statistic follows a chi-square distribution with one degree of freedom. When the d statistic exceeds the chosen critical value, we may reject the null hypothesis that these two functions are empirically equivalent.

(b) The data available for the study under consideration contains two blocks in each district and the random selection of circles, fields and plots are done with the blocks as the basis. Each year's data comprises of four blocks and hence the homogeneity of blocks over the years has been examined. The question arises whether the same regression relationship will apply to each block, when measurements have been made of several biometrical characters in different blocks of data. Even if the regression coefficients are equal for each block, the constant may differ, so that the regression lines will be parallel rather than coincident. If the regression equations for the different blocks are identical, even to the constant term, the blocks may be regarded as homogeneous in respect to the relationship of the dependent variable with the set of independent variables. Hence it follows that any difference between the mean values of the dependent variable for different blocks is attributable to differences in the mean values of the independent variables.

The significance of the difference among the regression coefficients for different blocks may be tested, by means of the analysis of variance. On the hypothesis that the regression coefficients in the population are the same, the common set of coefficients may be estimated from the combined sums of squares and products within the blocks, and the regressions sum of squares determined from the combined data. This sum of squares would be identical to the sum of the regression sums of squares from each block, if the regression coefficients were in fact the same for each set. Hence, the difference between the sum of the regression sums of squares for each block and the combined regression sum of squares gives a criterion appropriate for an overall test of differences among the coefficients.

To carry out these tests we shall require (i) the combined sums of squares and products for the different blocks, which we define as the sums of the corresponding sums of squares and products within blocks; and (iii) the overall sums of squares and products, which we define as the total sums of squares and products overall the blocks, regardless of block differences.

Suppose that there are m blocks of data, with p independent variables $x_1, x_2, \ldots, x_p$. We use the following notation:

n_r	number in the r^{th} block
n	total number
u_r	Sum of squares of y in the r^{th} block
p_{ri}	Sum of products with x_i in the r^{th} block
t_{rhi}	Sum of products of x_h and x_i in the r^{th} block
b_{ri}	partial regression coefficient on x_i in the r^{th} block
t_{chi}	Combined sum of products of x_h and x_i and, similarly, other quantities in the combined analysis will be indicated by the subscript c;
t_{ohi}	overall sum of products of x_h and x_i and similarly, other quantities in the overall analysis will be indicated by the subscript o.

(i) <u>To test differences among regression coefficients</u>

The sum of squares for regression for the r^{th} block is $\sum_i b_{ri}\, p_{ri}$, with p degress of freedom, where

$b_{ri} = \sum_h p_{rh}\ t_r^{hi}$, where t_r^{hi} is the quantity in the inverse matrix corresponding to the matrix (t_{rhi}).

The combined regression coefficients are $b_{ci} = \sum_h p_{ch}\ t_c^{hi}$, so that the sum of squares for the combined regression, also with p degress of freedom is $\sum_i b_{ci}\ \ p_{ci}$

Hence the sum of squares for testing difference of regressions is $\left\{ \sum_r \sum_i b_{ri} \; p_{ri} - \sum_i b_{ci} \; p_{ci} \right\}$

with (m-1)p degress of freedom. Since $p_{ci} = \sum_r (p_{ri})$,

the sum of squares of the above quantity can be written in the form $\sum_r \sum_i (b_{ri} - b_{ci}) \, p_{ri}$, in which its dependence on

the differences of the regression coefficients in the different blocks is clearly shown. The analysis of variance for testing difference of regressions is as follows:

Source	D.F.	Sum of Squares	Mean Square
Combined regression	p	$\sum_i b_{ci} \; p_{ci}$	
Difference of regressions	$(m-1)p$	$\sum_r \sum_i b_{ri} \; p_{ri} - \sum_i b_{ci} \; p_{ci}$	$\frac{1}{(m-1)p}\left[\sum\sum b_{ri} \; p_{ri} - \sum_i b_{ci} \; p_{ci}\right]$
Combined residual	$(n-mp-m)$	$u_c - \sum_r \sum_i b_{ri} \; p_{ri}$	$\frac{1}{n-mp-m}\left[u_c - \sum b_{ri} \; p_{ri}\right]$
Total within blocks	$n-m$	u_c	

(ii) Test of differences of position: If the difference of regressions is not significant, it may be assumed that the regression lines for the different blocks are parallel, the combined regression coefficients b_{ci} being applicable. The question then arises whether the differences in the position of these parallel lines are significant. To test this a single line is fitted to all the data, regardless of block differences. The overall regression coefficients are

$b_{oi} = \sum_h p_{oh} t_o^{hi}$ giving an overall regression sum of squares of $\sum_i b_{oi} p_{oi}$.

The overall sum of squares of y has n-1 degrees of freedom, an increase of m-1 degrees of freedom over the total within blocks, this increase representing the variation between blocks. The analysis of variance which includes tests of difference of regression and of position is as follows. Here care should be taken not to test difference of positions unless difference of regressions is in fact not significant.

Contd......18

Source	D.F.	Sum of Squares	Mean Square
Overall regression	p	$\sum_i b_{oi} \, p_{oi}$	
Difference of positions	$m-1$	$u_o - u_c - \sum_i b_{oi} \, p_{oi} + \sum_i b_{ci} \, p_{ci}$	$\frac{1}{m-1}\left[u_o - u_c - \sum_i (b_{oi} p_{oi} + b_{ci} \, p_{ci})\right]$
Difference of regressions	$(m-1)p$	$\sum_r \sum_i b_{ri} \, p_{ri} - \sum_i b_{ci} \, p_{ci}$	$\frac{1}{(m-1)p}\left[\sum\sum b_{ri} p_{ri} - \sum_i b_{ci} \, p_{ci}\right]$
Combined residual	$n-mp-m$	$u_c - \sum_r \sum_i b_{ri} \, p_{ri}$	$\frac{1}{(n-mp-m)}\left[u_c - \sum_r \sum_i b_{ri} \, p_{ri}\right]$
Total	$n-1$	u_o	

(c) The pilot studies were carried out in West Bengal State on Jute crop during the years 1970-71, 71-72, 72-73 and 73-74. For the purpose of the survey, a stratified two-stage random sampling design was adopted. Two districts and two community development blocks within each district, were covered under the survey during the years mentioned above. The community development blocks in the districts having large areas under the Jute crop were chosen for the study. Data were collected on the same set of blocks during all the four years. The totality of the Jute fields in these blocks constituted the population for the characters under study, with each of the blocks being regarded as a sub-population. The number of V.L.W/J.F.A. circles in the blocks were varying from six to seventeen. All the circles within each of the blocks were considered for the study and thus these circles constitute the strata. For each circle, a list of all the fields under special Jute package programme was prepared and a random sample of five fields was selected from each such list. In each of the fields selected, two plots of size each 2mx2m were located at random for recording detailed biometrical observations.

The fields within each circle (i.e. stratum) formed the first stage units and the plots within fields formed the second stage units. The two stage analysis was carried out to have information regarding the variation between fields within circles and between plots within fields. Also, between circles (i.e. between strata) variation was worked out for each block along with the above analysis and the results were pooled over blocks (with appropriate degrees of freedom) within each year, assuming homogeneity over blocks.

In the above two-stage sampling design the sampling variance of the mean is a function of the variance components at the two stages. In order to estimate the sampling error of the mean, it is necessary to obtain information on those components. Further, such information will also be useful to develop a suitable sampling scheme wherein the total sample is distributed in such a manner that maximum efficiency is attained.

For calculating the variance of mean of the characters under study, the following analysis of variance was carried out for each year, to separate the variations between circles, fields and plots within fields.

Let y_{ij}, be the measure of the character under study in the j^{th} plot of the i^{th} field of a stratum. The additive model is given as

$$y_{ij} = \mu + f_i + e_{ij}$$

where μ is general mean, f_i the effect of the i^{th} field and e_{ij} the effect of the j^{th} plot in the i^{th} field. Assuming that f_i and e_{ij} are uncorrelated and distributed with means zero and variance σ_f^2 and σ^2 respectively, the analysis of variance is given as below:

Analysis of Variance

Source of Variation	D.F.	M.S.	Expectation of M.S.
Between circles (i.e. strata)	$l-1$		
Between fields within circles	$\sum_{i=1}^{l} (m_i-1)$	s_f^2	$\sigma^2 + \lambda_1 \sigma_f^2$
Between plots within fields	$\sum_{i=1}^{l} \sum_{j=1}^{m_i} (n_{ij}-1)$	s^2	σ^2

where l is the number of circles, m_i is the number of fields in i^{th} circle, and n_{ij} is number of plots in j^{th} field of i^{th} circle.

Also

$$\lambda_1 = \frac{1}{\sum_{i=1}^{l} (m_i - 1)} \left[n_{..} - \sum_{i=1}^{l} \left\{ \frac{\sum_{j=1}^{m_i} n_{ij}^2}{n_{i.}} \right\} \right]$$

Where $n_{i.}$ is number of plots in the i^{th} circle and

$$n_{..} = \sum_{i=1}^{l} n_{i.}$$

The mean value worked out for a stratum is given as

$$\hat{\bar{y}} = \frac{\sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} y_{ijk}}{\sum_{j=1}^{m_i} n_{ij}}$$

Since the number of selected units at different stages did not vary much, instead of the exact formula for the variance of an estimated mean, the approximate formula given below was used:

$$V(\hat{\bar{y}}) = \frac{\sigma_f^2}{\bar{m}} + \frac{\sigma^2}{\bar{m}\,\bar{n}}$$

where $\bar{m}$ is the average number of fields in circles, $\bar{n}$ the average number of plots in fields and other notations are same as mentioned earlier. As the sampling fraction was samll at all stages the finite correction factors have been ignored.

A cost function

$$C = c_1 \bar{m} + c_2 \bar{m} \bar{n},$$

where c_1 and c_2 are the costs incurred during the survey to collect information for a field and a plot respectively, was considered so as to obtain the optimum values for $\bar{m}$ and $\bar{n}$. These were worked out by minimising the variance subject to fixed cost equal to ? C.

(d) The pilot studies which are under consideration aim at developing an appropriate regression equation so that yield can be forecast in advance of the harvest as accurately as possible. To forecast the yield in a year, the regression equation of its immediate past year has been utilised so far. Here the mean values of the biometrical characters (independent variables) are substituted in the equation developed with the help of previous data (i.e. retaining the same regression coefficients and the pure constant term) and the yield is estimated.

An attempt has been made to see whether the regression equation obtained by using the combined data of two or more preceeding years would improve the estimate of the yield compared to the one indicated earlier. Equations has been obtained using one previous year data, two previous years data and three previous years data and the estimates obtained thus by utilising the means of the biometrical characters, have been examined.

CHAPTER - III

RESULTS AND DISCUSSIONS

3.1 As has been informed in thé earlier chapters, the observations on biometrical characters were recorded on Jute crop periodically. Observations relating to the first period were recorded after four weeks of sowing of the crop and those for the subsequent periods at intervals of four weeks upto and including the time of harvest. The periods during which the observations [redacted] recorded were (inweeks) 3-7, 7-11, 11-15 and 15-19. The first two periods were too early in the sense that the plants will be in their initial stages of growth and hence the prediction of yield through the plant counts may be weak. The fourth and final period was close to harvest. Moreover in some cases of early harvest, observations on biometrical characters could not be obtained. As the preharvest estimates will be useful only if they are made reasonably in advance of the actual harvest, it was considered that the biometrical observations made at the third period would be the latest set which is likely to serve the purpose. Hence the results that follows were based on observations of third period only, which is considered as more appropriate than the other three periods.

The correlation matrices of yield y (i.e. fibre weight of Jute crop) with the biometrical characters (x_1 = plant density, x_2 = plant height and x_3 = basal diameter) for the years 1970-71 to 73-74 are presented in Table 3.1.

From the Table 3.1 it may be observed that the correlation coefficients between fibre weight and the number of plants were significant in all the years. Also correlation coefficients of yield with plant height were significant during the years 1970-71, 1971-72, and 1973-74. In case of Basal diameter the correlation coefficient was significant in the years 1971-72 and 1973-74. Thus fibre yield was correlated significantly in all the four years with no. of plants, in three out of four years with plant height and in two out of four years with Basal diameter.

3.2 (a) The multiple regression of fibre yield on the average number of plants per plot (of size four square metres), average height of the plant and average Basal diameter of the plant was worked out under the models mentioned in Chapter II 2.2(a). The observed values of the biometrical characters were subject to measurement errors. However, the errors were expected to be negligible because the measurement technique adopted during the field survey work had been standardized and the field staff were thoroughly trained in their use. The use of the classical regression theory treating the various biometrical observations as independent variates may be considered as valid.

By applying the four scalar transformations (linear, logarithmic, square root and reciprocal) to the three independent variables viz., plant density, plant height and Basal diameter, the 64 regression equations for each model for each year along with the corresponding multiple correlation coefficients have been presented in Tables 3.2.(a)(1) and 3.2.(a)(2) with y and log y as dependent variables respectively.

It is to be noted here that the equation which yields the highest value for the multiple correlation coefficient (R) will have the minimum residual sum of squares.

From the Table 3.2.(a)(1), it was observed that in respect of 43%, 83%, 36% and 47% of the equations during the years 1970-71, 71-72, 72-73 and 73-74 respectively, the multiple correlation coefficients values was less than that of the simple linear equations value (i.e. $y = a_1 + b_1 x_1 + c_1 x_2 + d_1 x_3$). In respect of the remaining equations for which the multiple correlation was higher than that for the simple linear equation, the largest difference in the R values was found to be 0.022, 0.006, 0.010 and 0.018 for the years 1970-71, 71-72, 72-73 and 73-74 respectively.

From the Table 3.2.(a)(2), it was observed that in respect of 59%, 75%, 92% and 66% of the equations during the years 1970-71, 71-72, 72-73 and 73-74 respectively, the multiple correlation coefficient (R) values was less than that of the log-linear equation's value (i.e. $\log y = a_2 + b_2 \log x_1 + c_2 \log x_2 + d_2 \log x_3$). In respect of the remaining equations for which the multiple correlation was higher than that for the log-linear equation, the largest difference in R values was found to be 0.043, 0.013, 0.002 and 0.018 for the years 1970-71, 71-72, 72-73 and 73-74 respectively.

It was observed that by transforming the independent variables in the equations, (keeping the dependent variable constant) the difference in the variability explained by these equations was only marginal. Hence from the practical point of convenience that it would appear that the simple linear and log-linear equations are more appropriate than the others.

To distinguish between these two equations a method has been suggested in Chapter II 2.2.(a). The two equations, simple linear and log-linear are denoted by (i) and (ii) and the corresponding transformed equations i.e. by multiplying the dependent variables by c, the inverse of the geometric mean of the y values, are denoted by (iii) and (iv). The values are presented in the Table 3.2(a)(3). From this table it was observed that for three out of four years equation (ii) is empirically a more appropriate functional form, by virtue of having the residual sum of squares smaller than the other. The year 1973-74 was an exception and in that year equation (i) was found appropriate. From the last column of the table, which gives the values d-statistic, for testing significance between residual sum of squares of the two functional forms, it was observed that the two functional forms were empirically equivalent for the years 1970-71, 72-73 and 73-74 and the log-linear equation was superior for the year 1971-72.

To sum up, the results indicate that practically both equations, linear and log-linear, are empirically equivalent except for one year's data. The log-linear equation is suggested as superior to the linear equation for its consistency of having higher value for the multiple correlation co-efficient.

3.2 (b) To test the homogeneity of regression equations over blocks for each of the years 1970-71, 71-72, 72-73 and 73-74, the method has been suggested in Chapter II 2.2.(b). The detailed analysis of variance was carried out for each year and the results are presented in Tables 3.2.(b)(1).

The results obtained in the tables show that the difference of regressions was not significant for all the four years. Hence it was found valid to use the combined regression over the blocks in each year. The difference of positions, were, of course, found to be highly significant for the years 1970-71, 71-72 and 72-73. In 1973-74 it was found to be non-significant. Thus the regression lines will be parallel and the distance between the lines gives the relative effects of the different blocks. Hence it was concluded that the blocks may be regarded as homogeneous in respect to the relationship of the dependent variable with the set of independent variables.

To test for the homogeneity of regression equations over years, the same method was applied and the results are presented in Table 3.2.(b)(2). Here also the differences among regression coefficients was observed to be non-significant while as the differences of positions was highly significant, which is so as variation exists over the years. Hence with respect to regression equations, it was found valid to use the combined regression over the years.

3.2 (c): A stratified two-stage random sampling design was adopted in the pilot studies as mentioned in the earlier chapters.

Two districts and two community development blocks within each district, were covered under the survey during all the four years. There were about six to seventeen V.L.W./J.F.A. circles in the blocks and these constitute the strata. Within each circle five fields were proposed to be selected at random and within each field two plots were located at random to record the measurements. Fields within circles formed the firststage units while as the second stage units were the demarcated plots within fields.

For studying the efficiency of stratification with the help of data from the pilot studies, the analysis of variance of plot yields were pooled over the different strata, assuming the number of circles are distributed among the strata in proportion to the crop areas and the order of magnitude of variation in the yield rate within different strata as the same. The details about such analysis of variance table has been mentioned in Chapter II 2.2(c). In some circles the number of fields selected were less than five rendering the data non-orthogonal. This fact has been taken into account and the appropriate formula has been used in working out estimates of true variance components. The analysis was carried out for each block and were pooled over the blocks i for each year assuming the homogeneity of blocks.

The Nested analysis of variance along with estimates of true variance components for each of the characters viz., fibre yield (y), plant density (x_1), plant height (x_2) and basal diameter (x_3) are presented in the tables 3.2.(c)(1).

With the help of these estimates $\hat{\sigma}_f^2$ and $\hat{\sigma}^2$ i.e. variance between fields within circles and variance between plots within fields respectively, the optimum values for $\bar{m}$ and $\bar{n}$, i.e. the average number of fields within each circle and the average number of plots within each field, were worked out for each of the biometrical character for all the four years data. This was done by minimising the variance

$$V = \left\{ \sigma_f^2 / \bar{m} + \sigma^2 / \bar{m}\,\bar{n} \right\}$$

subject to fixed cost $C = c_1 \bar{m} + c_2 \bar{m}\,\bar{n}$, where c_1 is the [redacted] cost attributed to field work only and c_2 to that of plot work only. The optimum values obtained in the process were

$$\bar{n}_{opt} = \sqrt{\frac{c_1}{c_2} \frac{\hat{\sigma}^2}{\hat{\sigma}_f^2}}$$

and $\bar{m}_{opt} = \dfrac{\hat{\sigma}_f^2 + \hat{\sigma}^2 / \bar{n}_{opt}}{V}$ or $\bar{m}_{opt} = \dfrac{C}{c_1 + c_2\, \bar{n}_{opt}}$

But the actual estimates of c_1 and c_2 values were not available. However, on the basis of empirical considerations based on enquiries it was felt that the ratios 1:1, 1:2 and 2:1 for $c_1 : c_2$ generally represent the situations covered. For these three ratios the values obtained for $\bar{m}$ and $\bar{n}$ are presented in Table 3.2.(c)(2).

optimum values of $\bar{m}$ and $\bar{n}$ after averaging out over the years, were as follows:

Ratio $c_1:c_2$		Plant density (x_1)	plant height (x_2)	Basal diameter (x_3)
1:1	$\bar{m}$	6	6	6
	$\bar{n}$	2	2	2
1:2	$\bar{m}$	8	6	6
	$\bar{n}$	1	1	2
2:1	$\bar{m}$	6	5	5
	$\bar{n}$	2	2	3

Hence from the results obtained, it was observed that the optimum values for the average number of fields per circle and the average number of plots per field was found to be generally 6 to 8 and 2 respectively for the three ratios considered.

3.2 (d): Using the combined data of two or three preceding years or of previous years to predict the current or future year's yield the following procedure was adopted. The average values of the biometrical characters were substituted in the fitted regression equation of the previous year (or combined previous years data) and the estimate of the yield, y was worked out. This estimate was compared with the observed value and the percentage difference was taken into consideration to draw the conclusions.

To predict the yield for the year 1973-74, we made use of the individual fitted reg ression equations of 1972-73, 1971-72, 1970-71; the combined regression of 1972-73, 1971-72 and 1972-73, 1971-72 and 1970-71. The same procedure was followed for the years 1972-73 and 1971-72. The results so obtained are tabulated in the Table 3.2(d)(1).

It was observed from the results obtained in Table 3.2.(d)(1) that for forecasting the yield of 1973-74, the estimate value obtained by making use of the regression equation of 1970-1971 was found to be very close to the observed value as the percentage difference being only 0.8. The regression equations made use of 1972-73 and 1971-72 explained the above difference to the extent of + 17.3 and -3.5 respectively. From the rainfall data for West Bengal during the periods April-May to October-November for the above four years, it was observed that the rainfall was more or less similar in case of 1973-74 and 1970-71 and variations were observed for the other two years. Hence in view of these variations of rainfall, the above results were justified since generally one express the difference should be small for the immediate past year to the year under consideration than when compared to the years which are much older. The same pattern of results were obtained for forecasting the yield of 1972-73 by making use of the fitted regression equations of 1971-72 and 1970-71. The percentage difference explained by the use of former equation was -21.91 and it was -15.74 with respect to the latter equation. It was of the order of + 43.7 when the fitted regression equation of 1970-1971 was used to forecast the yield of 1971-72. Hence it was observed that the use of the regression equation of the immediate past year, may not be strong unless the two years are similar with regard to factors like weather, rainfall etc.

Use of combined data of two or more years, stabilises the regressioncoefficients and also the seasonal variations will be taken into account and hence provides a better estimate than by using a single years data. Using two years data the percentage difference between the observed yield of 1973-74 and that estimated was 14.4 and it reduced to 7.9 when the data of 1970-71 was included to the above set of two years data.(Table 3.2.(d)(1)) The percentage difference was reduced from -21.9 to -20.4, though marginal, when the combined data of two years were used as against a single years data for forecasting the yield of 1972-73.

Hence in view of the above results, it was observed that the use of combined regression equation of the available previous years data was more appropriate than the use of any single year as it has got two advantages viz. the quantum of data being large and the seasonal variations between the years will taken care off.

SUMMARY

With a view to build up pre-harvest estimates of fibre yield of Jute crop on the basis of biometrical observations such as plant density, plant height and basal diameter of plants, recorded at the various growth stages of the crop, the Institute of Agricultural Research Statistics undertook pilot surveys in West Bengal State during the years 1970-71, 71-72, 72-73 and 73-74.

The survey was conducted in the districts of Nadia and 24-Parganas (N). Four community development blocks, two from each district viz. (i) Bhimpur (Nadia) (ii) Haringhatta (Nadia), (iii) Baduria (24-Parganas) and (iv) J.A.R.I. Extension (24-Parganas) were covered under the survey. The totality of these blocks constituted the population for the characters under study with each block being treated as a sub-population. The number of circles were varying from six to seventeen in the blocks totalling fifty in the four blocks together.

The design of the survey was stratified two-stage random sampling, with fields within circles as primary stage units and plots within fields as the secondary stage units. All the circles were covered under the survey and hence each of them constituted a stratum. From each circle, five fields growing Jute Crop were selected at random and within each field two plots each of size 5m x 2m were located at random for recording the detailed biometrical observations. The total number of plants were counted in each selected plot and the observations on height and basal

diameter were recorded on five plants (the four corner ones and the central plants). The first set of observations were recorded about a month after sowing of the crop; this was followed by periodic observations at intervals of four weeks upto and inclusive of the time of harvest.

The third period (i.e. observations made during the weeks 11-15 after sowing the crop) was considered as appropriate to forecast the yield in preferance to the other periods during which the observations were recorded.

The different transformations applied to the independent variables (i.e. on the three biometrical characters) in the prediction equations were found to be only of marginal improvement over the simple linear and Log-linear equations. Hence these two were considered to be empirically appropriate for prediction purposes. Between these two equations log-linear was found to be superior to the simple linear equation.

The combined regression equation over blocks within each year and over years were found to be valid as the regression coefficients were found to be homogeneous in either case.

For the sampling design adopted in the survey, the optimum number of fields per circle and plots per field were found to be 6 to 8 and 2 respectively.

To forecast the yield rate for the year 1973-74, the fitted regression equation based on the immediately preceding year was found to be inferior to the one based on the data pertaining to the years 1970-71 and 1971-72. This surprising result was attributed to the fluctuations in the weather, rainfall etc. over the years. However, the combined regression equation based on all the three previous years' data was found to yield better results. The combined equation has two advantages viz. a larger quantum of data and the averaging of seasonal variations over years.

Table 3.1: **Total correlation coefficents between yield and other biometrical characters during the years 1970-71 to 1973-74**

Year		Fibre yield of Jute (y)	No. of plants (x_1)	Plant Height (in mts) (x_2)	Basal diameter (in Cms) (x_3)
1970-71	y		0.43**	0.27**	0.07
	x_1			0.10	-0.06
	x_2				0.32
	x_3				
1971-72	y		0.33**	0.58**	0.21**
	x_1			0.12	-0.02
	x_2				0.68
	x_3				
1972-73	y		0.38**	0.05	-0.05
	x_1			-0.34	-0.45
	x_2				0.61
	x_3				
1973-74	y		0.17*	0.46**	0.28**
	x_1			-0.18	-0.37
	x_2				0.66
	x_3				

N.B: * and ** indicate signfficance at 5 per cent and 1 per cent level respectively.

Table 3.2.(a)(1) Multiple regression coefficients of fibre weight(y) on the biometrical observations (x_1,x_2,x_3) recorded at the third period of the growth of crop for the years 1970-71,71-72,72-73 and 73-74 under the following equations

S.No.	y	x_1	x_2	x_3	1970-71 R	1971-72 R	1972-73 R	1973-74 R
1	y	x_1	x_2	x_3	0.489	0.671	0.415	0.485
2				Log x_3	0.491	0.671	0.418	0.485
3				$\sqrt{x_3}$	0.490	0.672	0.416	0.485
4				$1/x_3$	0.493	0.666	0.420	0.486
5			Log x_2	x_3	0.482	0.643	0.417	0.479
6				Log x_3	0.484	0.647	0.419	0.479
7				$\sqrt{x_3}$	0.483	0.646	0.418	0.479
8				$1/x_3$	0.487	0.648	0.421	0.479
9			$\sqrt{x_2}$	x_3	0.486	0.661	0.416	0.483
10				Log x_3	0.488	0.663	0.419	0.483
11				$\sqrt{x_3}$	0.486	0.663	0.417	0.483
12				$1/x_3$	0.490	0.661	0.421	0.484
13			$1/x_2$	x_3	0.472	0.586	0.418	0.465
14				Log x_3	0.475	0.590	0.420	0.464
15				$\sqrt{x_3}$	0.473	0.588	0.419	0.465
16				$1/x_3$	0.479	0.594	0.422	0.464
17		Log x_1	x_2	x_3	0.508	0.677	0.414	0.493
18				Log x_3	0.509	0.677	0.417	0.493
19.				$\sqrt{x_3}$	0.508	0.677	0.416	0.493
20				$1/x_3$	0.511	0.671	0.418	0.494
21			Log x_2	x_3	0.500	0.648	0.414	0.487
22				Log x_3	0.502	0652	0.416	0.487

S. No.	y	x_1	x_2	x_3	1970-71 R	1971-72 R	1972-73 R	1973-74 R
23	y	Log x_1	Logx_2	$\sqrt{x_3}$	0.501	0.650	0.415	0.487
24				$1/x_3$	0.506	0.652	0.417	0.488
25			$\sqrt{x_2}$	x_3	0.504	0.666	0.415	0.491
26				Log x_3	0.506	0.668	0.416	0.491
27				$\sqrt{x_3}$	0.505	0.668	0.416	0.491
28				$1/x_3$	0.509	0.665	0.417	0.492
29			$1/x_2$	x_3	0.491	0.591	0.413	0.474
30				Log x_3	0.494	0.595	0.415	0.474
31				$\sqrt{x_3}$	0.493	0.593	0.414	0.474
32				$1/x_3$	0.499	0.599	0.516	0.473
33		$\sqrt{x_1}$	x_2	x_3	0.501	0.675	0.420	0.489
34				Log x_3	0.503	0.674	0.423	0.489
35				$\sqrt{x_3}$	0.502	0.675	0.421	0.489
36				$1/x_3$	0.505	0.669	0.425	0.490
37			Log x_2	x_3	0.494	0.647	0.421	0.483
38				Log x_3	0.496	0.651	0.423	0.483
39				$\sqrt{x_3}$	0.495	0.649	0.422	0.483
40				$1/x_3$	0.499	0.651	0.425	0.483
41			$\sqrt{x_2}$	x_3	0.498	0.664	0.420	0.487
42				Log x_3	0.500	0.666	0.423	0.487
43				$\sqrt{x_3}$	0.499	0.666	0.422	0.487
44				$1/x_3$	0.503	0.664	0.425	0.487
45			$1/x_2$	x_3	0.485	0.589	0.421	0.469
46				Log x_3	0.488	0.594	0.423	0.469
47				$\sqrt{x_3}$	0.486	0.592	0.422	0.469
48				$1/x_3$	0.492	0.598	0.424	0.468

S.No.	y	x_1	x_2	x_3	1970-71 R	1971-72 R	1972-7 R	
49	y	$1/x_1$	x_2	x_3	0.489	0.675	0.377	
50				Log x_3	0.490	0.674	0.378	
51				$\sqrt{x_3}$	0.489	0.675	0.377	
52				$1/x_3$	0.492	0.669	0.378	
53			Log x_2	x_3	0.482	0.644	0.376	[illegible]
54				Log x_3	0.483	0.648	0.377	[illegible]
55				$\sqrt{x_3}$	0.483	0.646	0.376	[illegible]
56				$1/x_3$	0.486	0.647	0.377	[illegible]
57			$\sqrt{x_2}$	$/x_3$	0.486	0.663	0.377	[illegible]
58				Log x_3	0.487	0.665	0.377	[illegible]
59				$\sqrt{x_3}$	0.486	0.665	0.377	[illegible]
60				$1/x_3$	0.489	0.662	0.378	[illegible]
61			$1/x_2$	x_3	0.473	0.583	0.374	[illegible]
62				Log x_3	0.475	0.588	0.376	[illegible]
63				$\sqrt{x_3}$	0.474	0.586	0.375	[illegible]
64				$1/x_3$	0.479	0.592	0.376	[illegible]

N.B: y is the average fibre weight of Jute crop in Kgs per plot of size 4 Sq. mts.

x_1 is the average number of plants per plot of size four Sq. mts.

x_2 is the average ~~[illegible]~~ height of the plants. (in metres)

x_3 is the average diameter of the plants (in centimetres).

Table: 3.2.(a)(ii) **Multiple regression coefficients of fibre weight (y) on the biometrical observations (x_1, x_2, x_3) recorded at third period of the growth of crop for the years 1970-71, 71-72, 72-73 and 73-74 under the following regression equations**

S.No.	y	x_1	x_2	x_3	1970-71 R	1971-72 R	1972-73 R	1973-74 R
1	Log y	x_1	x_2	x_3	0.517	0.700	0.429	0.498
2				Log x_3	0.519	0.699	0.432	0.498
3				$\sqrt{x_3}$	0.518	0.700	0.430	0.498
4				$1/x_3$	0.523	0.694	0.433	0.499
5			Log x_2	x_3	0.510	0.698	0.430	0.499
6				Log x_3	0.512	0.701	0.432	0.499
7				$\sqrt{x_3}$	0.511	0.700	0.432	0.499
8				$1/x_3$	0.516	0.700	0.430	0.500
9			$\sqrt{x_2}$	x_3	0.514	0.703	0.432	0.499
10				Log x_3	0.516	0.704	0.431	0.500
11				$\sqrt{x_3}$	0.515	0.704	0.431	0.500
12				$1/x_3$	0.519	0.701	0.433	0.501
13			$1/x_2$	x_3	0.498	0.662	0.431	0.490
14				Log x_3	0.502	0.6667	0.433	0.490
15				$\sqrt{x_3}$	0.500	0.664	0.432	0.490
16				$1/x_3$	0.507	0.669	0.433	0.490
17		Log x_1	x_2	x_3	0.586	0.715	0.459	0.511
18				Log x_3	0.588	0.714	0.461	0.512
19				$\sqrt{x_3}$	0.587	0.715	0.460	0.512
20				$1/x_3$	0.591	0.710	0.462	0.514
21			Log x_2	x_3	0.579	0.712	0.458	0.512
22				Log x_3	0.582	0.715	0.460	0.512
23				$\sqrt{x_3}$	0.580	0.714	0.462	0.512

S.No.	y	x_1	x_2	x_3	1970-71 R	1971-72 R	1972-73 R	1973-74 R
24	Log y	Log x_1	Log x_2	$1/x_3$	0.585	0.714	0.461	0.513
25			$\sqrt{x_2}$	x_3	0.583	0.717	0.458	0.512
26				Log x_3	0.585	0.718	0.460	0.513
27				$\sqrt{x_3}$	0.584	0.718	0.460	0.513
28				$1/x_3$	0.588	0.715	0.459	0.514
29			$1/x_2$	x_3	0.570	0.677	0.457	0.504
30				Log x_3	0.574	0.681	0.458	0.504
31				$\sqrt{x_3}$	0.571	0.679	0.459	0.504
32				$1/x_3$	0.578	0.684	0.458	0.504
33		$\sqrt{x_1}$	x_2	x_3	0.553	0.708	0.448	0.504
34				Log x_3	0.555	0.707	0.451	0.505
35				$\sqrt{x_3}$	0.554	0.708	0.450	0.504
36				$1/x_3$	0.558	0.702	0.451	0.506
37			Log x_2	x_3	0.545	0.706	0.449	0.505
38				Log x_3	0.548	0.708	0.450	0.505
39				$\sqrt{x_3}$	0.547	0.707	0.450	0.505
40				$1/x_3$	0.552	0.708	0.451	0.506
41			$\sqrt{x_2}$	x_3	0.549	0.710	0.449	0.505
42				Log x_3	0.552	0.711	0.451	0.506
43				$\sqrt{x_3}$	0.551	0.711	0.450	0.506
44				$1/x_3$	0.555	0.708	0.451	0.507
45			$1/x_2$	x_3	0.535	0.670	0.448	0.497
46				Log x_3	0.539	0.674	0.449	0.496
47				$\sqrt{x_3}$	0.537	0.672	0.449	0.497
48				$1/x_3$	0.544	0.677	0.449	0.496
49		$1/x_1$	x_2	x_3	0.621	0.726	0.457	0.527

S.No.	y	x_1	x_2	x_3	1970-71 R	1971-72 R	1972-73 R	1973-74 R
50	Log y	$1/x_1$	x_2	Log x_3	0.622	0.725	0.458	0.528
51				$\sqrt{x_3}$	0.621	0.726	0.458	0.527
52				$1/x_3$	0.625	0.721	0.457	0.530
53			Log x_2	x_3	0.614	0.721	0.456	0.527
54				Log x_3	0.617	0.724	0.456	0.528
55				$\sqrt{x_3}$	0.615	0.723	0.456	0.527
56				$1/x_3$	0.620	0.723	0.455	0.528
57			$\sqrt{x_2}$	x_3	0.618	0.727	0.457	0.528
58				Log x_3	0.620	0.728	0.457	0.529
59				$\sqrt{x_3}$	0.619	0.728	0.457	0.528
60				$1/x_3$	0.622	0.725	0.456	0.530
61			$1/x_2$	x_3	0.606	0.684	0.454	0.520
62				Log x_3	0.609	0.688	0.454	0.520
63				$\sqrt{x_3}$	0.608	0.686	0.454	0.520
64				$1/x_3$	0.613	0.691	0.453	0.520

N.B.: y is the average fibre weight of Jute crop in kgs per plot of size four square metres.

x_1 is the average number of plants per plot of size four square metres.

x_2 is the average height of plants in metres.

x_3 is the average diameter of plants in centimetres.

Table: 3.2 (a)(3)

Years	c-value	Equations	Residual Sum of Squares	d-statistic (computed)
1970-71	1.1941	(i) (ii) (iii) (iv)	21.8994 21.1447 21.8077 21.1447	1.35
1971-72	2.1710	(i) (ii) (iii) (iv)	8.2353 31.6505 38.8110 31.6505	11.37**
1972-73	0.9979	(i) (ii) (iii) (iv)	25.2403 22.8452 25.1348 22.8452	4.65
1973-74	1.1105	(i) (ii) (iii) (iv)	9.3075 13.0258 11.4786 13.0258	6.05

N.B. ** indicate significance at 1 per cent level

Table 3.2(b)(1)

Test of differences among regression coefficients and of positions over blocks

Years	Source	D.F.	Sum of Squares	Mean Square
1970-71	Overall regression	3	6.8803	
	Difference of positions	3	3.2393	1.0798**
	Difference of regressions	9	1.1581	0.1297(n)
	Combined residual	186	17.5020	0.0941
	Total	202	28.7797	
1971-72	Overall regression	3	6.7607	0.
	Difference of positions	3	1.3534	0.4511**
	Difference of regressions	9	0.2594	0.0288(n)
	Combined residual	241	6.6226	0.0279
	Total	256	14.9961	
1972-73	Overall regression	3	4.1194	
	Difference of positions	3	2.9902	0.9967**
	Difference of regressions	9	1.2506	0.1389(n)
	Combined residual	208	20.9995	0.1009
	Total	223	29.3597	
1973-74	Overall regression	3	3.0847	
	Difference of positions	3	0.0969	0.0323(n)
	Difference of regressions	9	0.7388	0.0821(n)
	Combined residual	204	9.2369	0.0453
	Total	249	13.1583	

N.B: (n) Not significant
** Significant at 1 per cent level

Table 3.2(b)(2)

Test of differences among regression coefficients and of positions over the years

Source	D.F.	Sum of Squares	Mean Square
Overall regression	3	45.1108	
Difference of positions	3	13.5646	4.4912**
Difference of regressions	9	0.8393	0.0938 (n)
Combined residual	887	68.4477	0.0772
Total	902	127.9624	

N.B. (n) not significant

** significant at 1 per cent level

Table 3.2 (c)(1)

Nested analysis of variance with true variance components for the year 1970-71

Character	Source of Variation	D.F.	M.S.	Variance components
y	Between Circles	43	0.6188	
	Between fields within circles	155	0.1303	0.0432
	Between plots within fields	202	0.0184	0.0184
x_1	Between Circles	43	7714.1550	
	Between fields within circles	155	2337.5367	633.0571
	Between plots within fields	202	695.7029	695.7029
x_2	Between Circles	43	0.8871	
	Between fields within circles	155	0.2741	0.0859
	Between plots within fields	202	0.0515	0.0515
x_3	Between Circles	43	0.5344	
	Between fields within circles	155	0.0725	0.0151
	Between plots within fields	202	0.0332	0.0332

Nested analysis of variance with true variance components for the year 1971-72

Character	Source of Variation	D.F.	M.S.	Variance components
y	Between Circles	48	0.1636	
	Between fields within circles	205	0.1026	0.0371
	Between plots within fields	257	0.0100	0.0100
x_1	Between Circles	48	4248.1770	
	Between fields within circles	205	3068.4809	841.4700
	Between plots within fields	257	966.8579	966.8579
x_2	Between Circles	48	0.3249	
	Between fields within circles	205	0.2816	0.0976
	Between plots within fields	257	0.0374	0.0374
x_3	Between Circles	48	0.2253	
	Between fields within circles	205	0.1186	0.0383
	Between plots within fields	257	0.0230	0.0230

Nested analysis of variance with true variance components for the year 1972-73

Character	Source of Variation	D.F.	M.S.	Variance components
y	Between Circles	46	0.5629	
	Between fields within circles	174	0.1315	0.0435
	Between plots within fields	224	0.0195	0.0195
x_1	Between Circles	46	6413.9970	
	Between fields within circles	174	2767.6128	781.4218
	Between plots within fields	224	756.7488	756.7488
x_2	Between Circles B	46	1.1849	
	Between fields within circles	174	0.2964	0.0823
	Between plots within fields	224	0.0776	0.0776
x_3	Between Circles	46	0.1541	
	Between fields within circles	174	0.0696	0.0154
	Between plots within fields	224	0.0302	0.0302

Nested analysis of variance with true variance components for the year 1973-74

Character	Source of Variation	D.F.	M.S.	Variance Components
y	Between circles	44	0.1778	
	Between fields within circles	171	0.0975	0.0290
	Between plots within fields	220	0.0232	0.0232
x_1	Between circles	44	6195.6068	
	Between fields within circles	171	1978.0309	453.3987
	Between plots within fields	220	816.6613	816.6613
x_2	Between Circles	44	0.9322	
	Between fields within circles	171	0.2025	0.0556
	Between Plots within fields	220	0.0600	0.0600
x_3	Between Circles	44	0.2874	
	Between fields within circles	171	0.0479	0.0079
	Between Plots within fields	220	0.0278	0.0278

Table 3.2.(c)(2)

Optimum values for the average number of fields within circles $(\bar{m})$ and average number of plots within fields $(\bar{n})$ with percentage standard error of the estimates of 10.5, 5.8 and 6.3 for the three biometrical characters respectively.

Ratio $c_1:c_2$	Years		Plant density (x_1)	Plant height (x_2)	Basal diameter (x_3)
	1970-71	$\bar{m}$	6	6	5
		$\bar{n}$	2	1	2
1:1	1971-72	$\bar{m}$	7	6	9
		$\bar{n}$	2	1	1
	1972-73	$\bar{m}$	8	7	5
		$\bar{n}$	1	1	2
	1973-74	$\bar{m}$	5	4	4
		$\bar{n}$	2	2	2
1:2	1970-71	$\bar{m}$	7	6	5
		$\bar{n}$	1	1	2
	1971-72	$\bar{m}$	10	6	9
		$\bar{n}$	1	1	1
	1972-73	$\bar{m}$	8	7	7
		$\bar{n}$	1	1	1
	1973-74	$\bar{m}$	7	5	4
		$\bar{n}$	1	1	2
2:1	1970-71	$\bar{m}$	6	5	4
		$\bar{n}$	2	2	3
	1971-72	$\bar{m}$	7	6	8
		$\bar{n}$	2	1	2
	1972-73	$\bar{m}$	6	5	5
		$\bar{n}$	2	2	2
	1973-74	$\bar{m}$	5	[illegible]	[illegible]
		$\bar{n}$	2		

Table 3.2.(d)(1)

Percentage difference between the observed value for yield during 1973-74 with those estimated from fitted regression equations of previous years data and those of combined data of two and three previous years

Fitted regression equation used of	Observed yield (1973-74) (y)	Estimated yield (y)	Percentage difference
1972-73	0.9349	1.0960	+ 17.3
1971-72	0.9349	0.9021	- 3.5
1970-71	0.9349	0.9418	+ 0.8
1972-73 + 1971-72	0.9349	1.0693	+ 14.4
1972-73 + 1971-72 + 1970-71	0.9349	1.0094	+ 7.9

N.B: y is the average fibre weight of Jute crop in Kg per plot of size 4 sq. mt.

Table 3.2.(d)(1)

Percentage difference between the observed value for yield during 1972-73 with those estimated from fitted regression equations of previous years data and those of combined data of two previous years

Fitted regression equation used of	Observed value (1972-73) (y)	Estimated value (y)	Percentage difference
1971-72	1.0604	0.8277	- 21.91
1970-71	1.0604	0.8931	- 15.74
1971-72 + 70-71	1.0604	0.8441	- 20.36

Percentage difference between the observed value for yield during 1971-72 with that estimated from fitted regression equation of 1970-71

Fitted regression equation used of	Observed value (1971-72) (y)	Estimated value (y)	Percentage difference
1970-71	0.5160	0.7418	+ 43.7

N.B: y is the average fibre weight of Jute crop in Kg per plot of size four sq. mets.

REFERENCES

1. Cochran, W.G. (1963). "Sampling Techniques" New York; John Wiley & Sons.

2. Draper, N.R. and Smith H. (1966). "Applied Regression Analysis" New York; John Wiley & Sons.

3. Johnston, J. (1963). "Econometric Methods" New York; McGraw-Hill Book Company.

4. Rao, P. and Miller, R.L. (1972). "Applied Econometrics" New Delhi; Prentice-Hall of India Private Limited.

5. Sanderson, Fred H. (1954). "Methods of Crop Forecasting" Cambridge Massachusetts; Harvard University Press.

6. Williams, E.J. (1959). "Regression Analysis" New York; John Wiley & Sons.

7. Box, G.E.P. and Cox, D.R. (1964) "An Analysis of Transe formations" Journal of the Royal Statistical Society, Series B.